

СТРАТЕГІЇ ПІДГОТОВКИ ФАХІВЦІВ У ГАЛУЗІ АНАЛІЗУ ДАНИХ

Кам'янець-Подільський національний університет імені Івана Огієнка

Анотація

Пропонується груповий метод виконання навчальних проєктів комп'ютерної обробки інформаційних даних з метою забезпечення професійних навиків та фахових компетентностей фахівців у галузі аналізу даних.

Ключові слова: аналіз великих даних, аналітик даних, інженер даних, дослідник даних.

Abstract

A group method of performing educational projects of computer processing of information data is proposed in order to ensure the professional skills and professional competences of specialists in the field of data analysis.

Keywords: big data analysis, data analyst, data engineer, data researcher.

Вступ

Об'єктом вивчення для здобувачів вищої освіти другого (магістерського) рівня вищої освіти, що навчаються на спеціальності 122 Комп'ютерні науки за освітньо-професійною програмою «Комп'ютерні науки та інформаційні технології» є процеси збору, представлення, обробки, зберігання, передачі та доступу до інформації в комп'ютерних системах. Отже, для них матеріалом глибокого аналізу є процеси, пов'язані з даними.

Про важливість аналізу інформаційних даних досить давно (більше двох століть тому) наголошував англійський банкір, бізнесмен і фінансист Натан Ротшильд, заявивши «Хто володіє інформацією – той володіє світом». Хоча крилатим цей вислів став лише після того, як його процитував Вінстон Черчіль, популярність та успіх дотримання цього постулату відчувалася і відчувається досі в усіх сферах людської діяльності, зачіпаючи інтереси мільйонів представників абсолютно різноманітних професій. У зв'язу з цим постає питання про важливість ефективної підготовки професіоналів у галузі обробки даних.

Вимоги роботодавців

Перш за все будемо виходити з вимог, що їх ставлять до фахівців з комп'ютерної аналітики даних.

У сучасному інформаційному просторі чітко виділяються три категорії спеціалістів у сфері обробки інформації. Мається на увазі аналітик даних, інженер даних і дослідник даних (data analyst, data engineer і data scientist) [1, 2]. Звичайно, це загальна класифікація. В залежності від функціональних обов'язків виокремлюють вузько кваліфікованих фахівців, наприклад, аналітиків-теоретиків, бізнес-аналітиків, фінансових аналітиків, соціоаналітиків, вебаналітиків, рекламних аналітиків, менеджерів тощо.

Їх знання з теорії даних застосовуються в широкому діапазоні професій і формуються на міждисциплінарній науці про дані, що базується на науково обґрунтованих методах, алгоритмах та принципах для збору або виділення інформації та ідей з різного типу даних (нечітких, структурованих чи неструктурованих), в тому числі великих.

Ці знання про дані дослідники отримують із банку даних за допомогою сучасних експертних методів, високопродуктивних алгоритмів, систем та інструментів інтелектуального аналізу даних. Накопичений досвід дозволяє отримувати інформації як з хороших (структурованих) даних, які є досить зручними для зберігання в наперед визначеному форматі, так і з нечітких, неструктурованих даних, які містять інформацію різного роду та зберігаються у широкому розмаїті форматів.

Зазначені вище три категорії спеціалістів мають різні призначення у сфері аналізу даних зі своїми особливостями та зоною відповідальності.

Зокрема, аналітики даних забезпечують раціональні методи збору даних, їх структурування, систематизацію, статистичний аналіз, в якійсь мірі, візуалізацію даних, відшуковують залежності і

формують прогнози на основі даних. Це дає їм змогу ухвалювати виважені рішення на основі аналізу даних.

Найуживанішими математичними методами, які використовуються ними в аналізі даних, є статистичні методи, а для успішної, презентабельної презентації своїх дієвих рекомендацій аналітик даних повинен добре розумітися в методах та інструментах візуалізації. Щоб рішення були переконливішими йому потрібно, м'яко кажучи, продумано подавати свій продукт у вигляді презентації.

Отже, аналітики даних повинні володіти відповідними компетентностями з баз даних, математичної статистики, веб-дизайну.

Інженера даних, відверто кажучи, не цікавить конкретний зміст інформації, яка міститься в даних і практично усі компетентності, що є важливими для аналітика даних, йому не актуальні. Він відповідає за розробку та підтримку інфраструктури зберігання й обробки даних, розробляє основу для різних операцій з даними, відповідає за розробку формату, над яким працюватимуть дослідники та аналітики даних, будує і керує системами збору, зберігання та обробки даних, розробляє і втілює рішення щодо баз даних, процесів та інших інструментів для ефективної роботи з великими обсягами даних. Виходячи із такої багатогранності обов'язків, мабуть варто говорити про колектив інженерів. Зрозуміло, що вони мають працювати як зі структурованими, так і з неструктурованими даними. Тому їм також потрібен досвід роботи з базами даних. Інженери даних орієнтуються в основних концепціях і алгоритмах програмування, добре розуміються на інструментах проектування та тестування.

Окрім навиків роботи з базами даних для інженера даних роботодавці зацікавлені у наявності навиків з програмування та вміння працювати колективно.

Дослідник даних займається практичним використанням даних, розробленням моделей прогнозування та розв'язання складних проблем. Він використовує методи машинного пошуку знань, статистичні методи та алгоритми аналізу даних, методи нечіткої математики, розробляють моделі і прогнозують алгоритмічні результати. З позицій роботодавців їм потрібні спеціалізовані фахівці, які володіють знаннями в галузі статистичних інструментів і навичками програмування обробки даних, знаннями алгоритмів машинного навчання, штучного інтелекту.

Шляхи організації навчання фахівців

Виходячи з вимог роботодавців, для належної підготовки фахівців у галузі аналізу даних необхідно забезпечити здобувачів вищої освіти фундаментальними знаннями сучасної науки про дані, яку можна розглядати як симбіоз наукових напрямів, що охоплює всі операції з даними, як-от методи отримання даних, зберігання та обробка даних, аналіз даних і прогнозування даних для отримання необхідної інформації. Це специфічна галузь, яка має спільне коріння з математикою, статистикою та комп'ютерним програмуванням. На гармонійне поєднання цих «трьох китів» потрібно орієнтуватися при організації освітнього процесу ІТ-фахівців з аналізу даних.

В теоретичному плані для здобувачів вищої освіти рівня магістр математичний апарат давно сформований і зосередитись потрібно на сучасних проблемах, пов'язаних, наприклад, з опрацюванням даних великих розмірів, нечіткою структурою даних тощо. Статистичні проблеми також озвучені при підготовці бакалаврів і тепер доцільно направити теоретичні дослідження на надійність прогнозів. В свою чергу, практика освоєння курсу програмування підказує, що комп'ютерне програмування перш за все повинно зосередитись на питаннях боротьби з шумами та організацією високопродуктивних обчислень.

Найбільш вузьким місцем виступає практика організації лабораторних занять. Для їх проведення, мабуть, не доцільно застосовувати традиційні індивідуальні завдання. Хоча вони також є і залишаються досить дієвими. Краще підходить груповий (колективний) метод. Такий висновок мотивований перспективою працевлаштування у фірмах з великою кількістю мікро-колективів, але не можна при цьому забувати, що значна кількість фахівців виконує вузькі індивідуальні завдання.

Груповий метод навчання передбачає організацію освітнього процесу, за якої професійні навички набуваються в процесі спілкування між здобувачами (взаємонавчання) у групах. Найважче – підібрати проєктні завдання. Вхідну інформацію економічного характеру одержати на фірмах, зазвичай, досить складно. Потрібно ретельно попрацювати над матеріалами виробничих практик, переглянути репозиторій з кваліфікаційними роботами здобувачів економічних спеціальностей, шукати можливості виконання кваліфікаційних робіт на замовлення. Серед інших джерел цікавим є статистичний матеріал освітнього процесу на факультеті. Можна назвати і ряд інших джерел. Загалом кількість групових

проектів не повинна перевищувати п'ять-шість. Інакше не вдасться забезпечити дієвий контроль в організації заняття.

Виходячи з цього, група може складатися з трьох і більше здобувачів, і не обов'язково повинна бути постійною. Роль кожного члена в групі розподіляють здобувачі самостійно. Перше організаційне заняття проводимо колективно, де кожен здобувач виставляє рейтинги запропонованих проектів, розробляє стратегічний перспективний та тактичний щоденний план. На основі рейтингу формуються команди, які обговорюють плани учасників групи і приступають до виконання проектів.

Така організація освітнього процесу, на наш погляд, має ряд позитивних моментів, зокрема, в процесі реалізації такої підготовки формуються навички проведення наукового дослідження та його оформлення, навички пошуку, використання та опрацювання інформації з різних джерел, чітко виділяються задачі кожної з трьох категорій спеціалістів у сфері обробки інформації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ірина Чиж. Data analyst, data engineer и data scientist – в чому різниця? [Електронний ресурс]. – Режим доступу: <https://dan-it.com.ua/uk/blog/data-analyst-data-engineer-i-data-scientist-v-chem-raznica/>
2. Data Science проти Data Mining: ключові відмінності. [Електронний ресурс]. – Режим доступу: <https://www.unite.ai/uk/data-science-vs-data-mining-key-differences/>

Мястковська Марина Олександрівна – кандидат педагогічних наук, старший викладач кафедри комп'ютерних наук, Кам'янець-Подільський національний університет імені Івана Огієнка, м. Кам'янець-Подільський, marinenka1@gmail.com.

Щирба Віктор Самуїлович – кандидат фізико-математичних наук, доцент, доцент кафедри комп'ютерних наук, Кам'янець-Подільський національний університет імені Івана Огієнка, м. Кам'янець-Подільський.

Maryna Oleksandrivna Miastkovska – Candidate of Pedagogical Sciences, senior lecturer of the Department of Computer Science, Kamianets-Podilskyi Ivan Ohiienko National University, Kamianets-Podilskyi, marinenka1@gmail.com.

Shchyryba Viktor Samuilovych – Candidate of Physical and Mathematical Sciences, Associate Professor, Associate Professor of the Department of Computer Science, Kamianets-Podilskyi Ivan Ohiienko National University, Kamianets-Podilskyi.